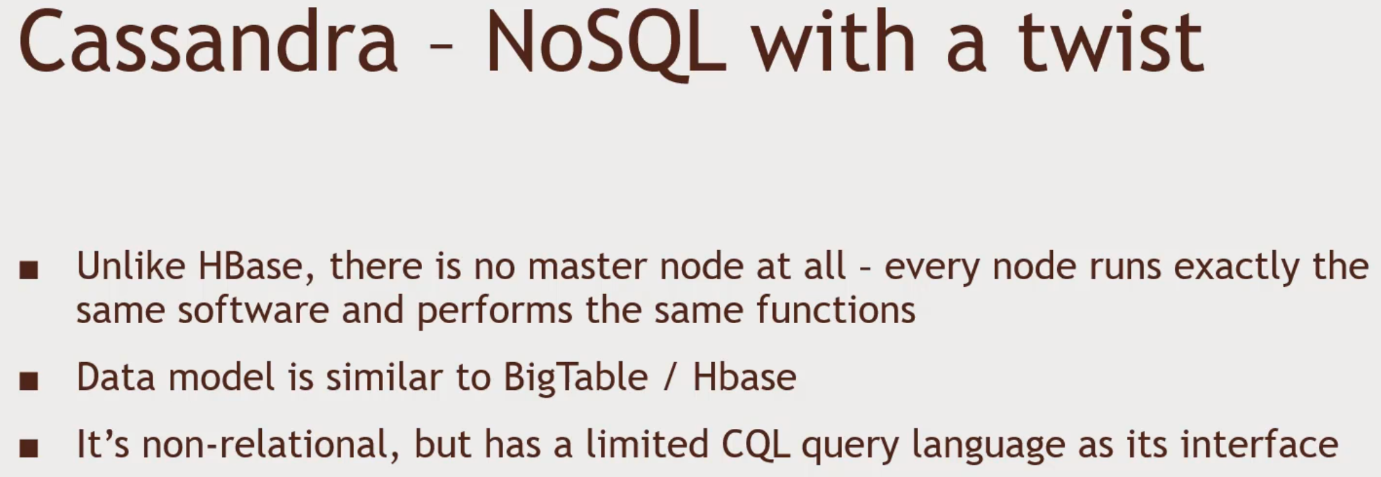
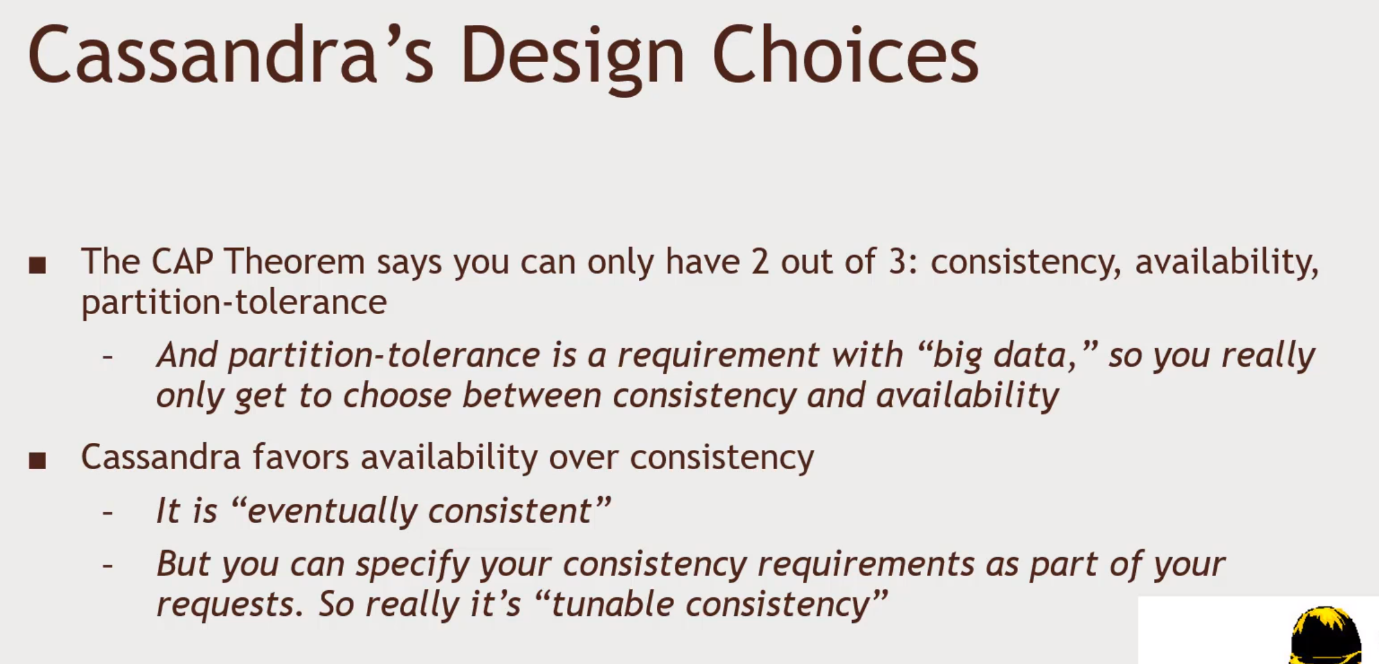
Cassandra Distributed database with no single point of failure. No master nodes

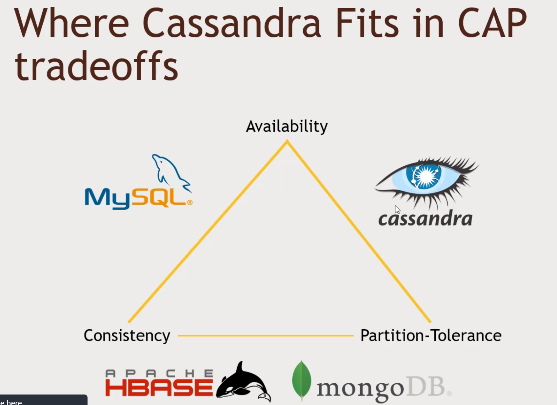
CQL query language.

Why Cassandra?

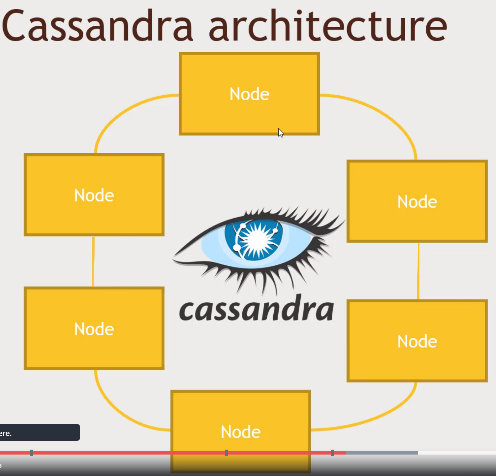
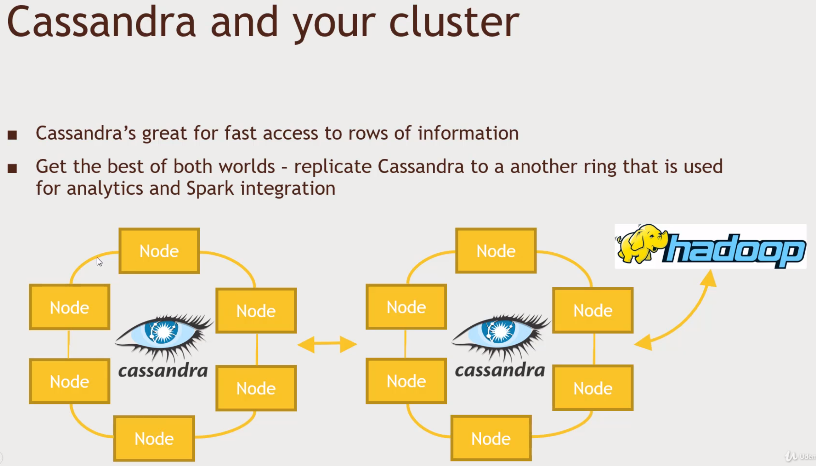
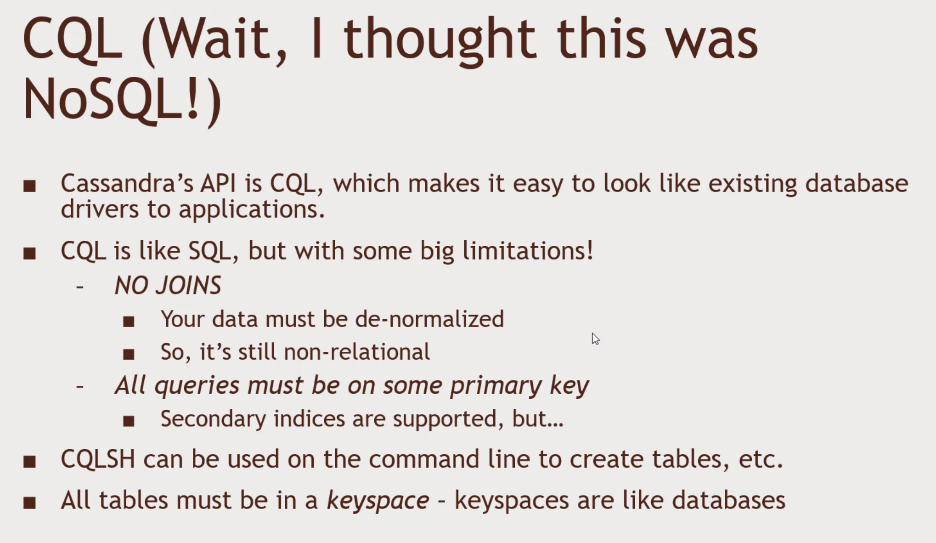
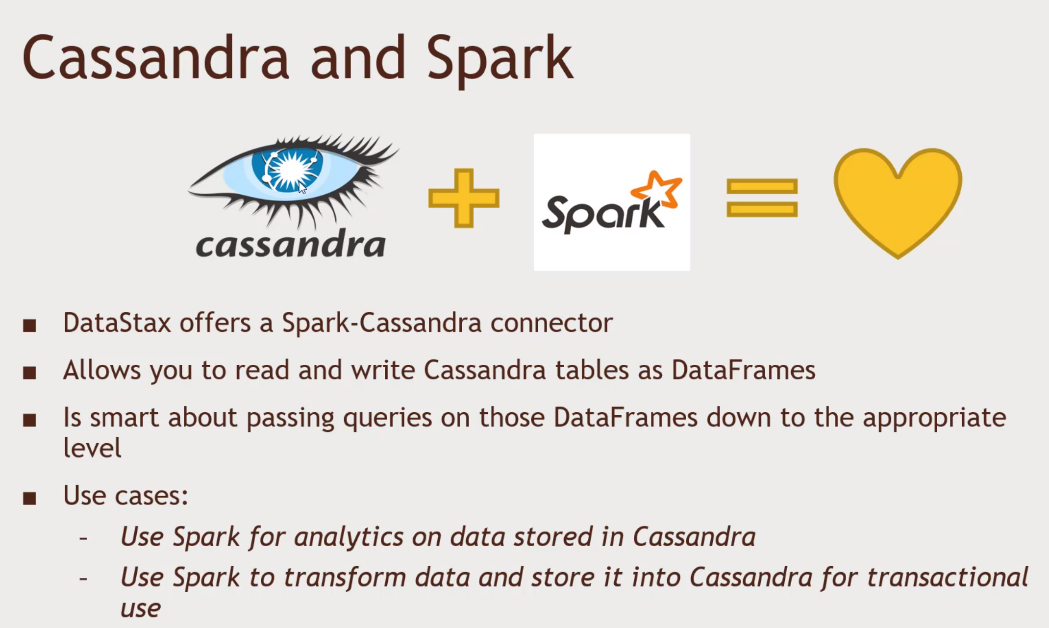
Consistency - write something to data base get the results no matter what.

Availability – Data base to be reliable,

Partition – split up and partitioned

Tunable consistency – more nodes to come back with the required answer

Hbase – favor in consistency. Instead of availability.

Cassandra have the gossip nature. Gossip Protocol. Each node is talking to other nodes. Auntie like structure. Distributed data among lots of nodes. Data get hashed into different nodes. The nodes talks to each other, peer to peer and etc.

Install Cassandra in HortonWorks – Python version 2.7 needed. Install both python 2.6 and 2.7

By yum install scl-utils.

Yum install centos-release-scl-rh

Yum install python27

Scl enable python27 bash

Python -v, run python 2.7

Cd /etc/yum.repos.d

Create vi datastax.repo, press I for insert and type the following:

Name = DataStax Repo for Apache Cassandra

Baseurl = <http://rpm.datastax.com/community>

Enable = 1

Gpgcheck = 0, then esc, :wq to quit and cat datstax.repo to check

To install Cassandra – yum install dsc30

Install python dependency – pip install cqlsh

Then service Cassandra start

Then cqlsh

To create a table in cqlsh > CREATE KEYSPACE movielens WITH replication = {‘class’: ‘SimpleStrategy’, ‘replication\_factor’ : ‘1’} AND durable\_writes = true;

USE movielens;

Prompt will show cqlsh:movielens> CREATE TABLE users (user\_id int, age int, gender text, occupation text, zip text, PRIMARY KEY (user\_id));

SELECT \* FROM users; will show nothing.

Write a Spark script to write data into Cassandra

Wget <http://media.sundog-soft.com/hadoop/CassandraSpark.py>

Check in ambari to see if the file exists.

Run the CassandraSpark.py.

In Putty: spark-submit –packages datastax:spark-cassandra-connector:2.0.0-M2-s\_2.11 CassandraSpark.py

Cqlsh> USE movieles;

Cqlsh:movielens> SELECT \* FROM users LIMIT 10;

Stop service: service cassandra stop